

Synchronization Properties of Network Motifs

I. LODATO¹, S. BOCCALETTI², AND V. LATORA³

¹ *Scuola Superiore di Catania, Via S. Paolo 73, 95123 Catania, Italy*

² *CNR- Istituto dei Sistemi Complessi, Via Madonna del Piano, 10, 50019 Sesto Fiorentino (FI), Italy*

³ *Dipartimento di Fisica e Astronomia, Università di Catania, and INFN, Via S. Sofia 64, 95123 Catania, Italy*

PACS. 89.75.-k - .

PACS. 05.45.Xt - .

PACS. 87.18.Sn - .

Abstract. – We address the problem of understanding the variable abundance of 3-node and 4-node subgraphs (motifs) in complex networks from a dynamical point of view. As a criterion in the determination of the functional significance of a n -node subgraph, we propose an analytic method to measure the stability of the synchronous state (SSS) the subgraph displays. We show that, for undirected graphs, the SSS is correlated with the relative abundance, while in directed graphs the correlation exists only for some specific motifs.

Recent empirical evidences indicate that complex networks, among other common properties, are characterized by the presence of various length cycles and specific motifs [1–3]. A motif M is a pattern of interconnections occurring either in a undirected or in a directed graph G at a number significantly higher than in randomized versions of the graph, i.e. in graphs with the same number of nodes and links (and eventually degree distribution) as the original one, but where the links are distributed at random. As a pattern of interconnections, is usually meant a small connected (undirected or directed) n -node graph M which is a subgraph of G . The concept of motifs was originally introduced by U. Alon and coworkers, who studied motifs in biological and non-biological networks [4–8]. The research of the significant motifs in a graph G is based on matching algorithms counting the total number of occurrences of each n -node subgraph M in the original graph and in the randomized ones. The statistical significance of M is then described by the Z -score, defined as: $Z_M = \frac{\#_M - \langle \#_M^{rand} \rangle}{\sigma_{\#_M}^{rand}}$, where $\#_M$ is the number of times the subgraph M appears in G , and $\langle \#_M^{rand} \rangle$ and $\sigma_{\#_M}^{rand}$ are, respectively, the mean and standard deviation of the number of appearances in the randomized network ensemble [5, 7]. The reasons of the variable frequency of different n -node subgraphs in a specific network are still poorly understood. There are at least two possible explanations. On the one hand, it is possible that certain constraints on the growth mechanism of a network as a whole determine which motifs become abundant [9, 10]. On the other hand, it is well known that the structure has important consequences on the network dynamics and functional robustness. So

that a particular n -node graph can become overrepresented because, due to its structure, it possesses some relevant functional properties [5].

In this letter, we address the question of network motifs in biological networks from a dynamic systems point of view. Naturally, a comprehensive analysis of the dynamics of networks is considerably more complicated than the corresponding analysis of their structure. This is due to the potentially complex functional dependencies between nodes, and to lack of knowledge of the specific interaction parameters. For such a reason, instead of modeling in details one particular biological network, we analyze the generic dynamic properties that arise from the topology of a n -node graph. In particular, we focus on the emergence of collective dynamic behaviors, such as *synchronization*, that is relevant in many biological systems, and we propose an analytic method to estimate the stability of the synchronous state (SSS) displayed by a n -node graph. We finally show that the SSS, potentially, can help explaining why certain network motifs are overrepresented in some real biological networks, while others are not.

We assume that the dynamics of a n -node motif M can be represented as a system of n ODE's:

$$\dot{\mathbf{x}}_i = \mathbf{f}_i(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \quad i = 1, \dots, n \quad (1)$$

where $\mathbf{x}_i \in R^m$ is the m -dimensional vector describing the state of node i (for instance the concentration of molecule i in a metabolic reaction, or the polarization state of neuron i in a neural network), and $\mathbf{f}_i : R^{m \times n} \rightarrow R^m$ is the function representing the effects on \mathbf{x}_i of all the nodes connected to i . In particular, we are neglecting the influence of the other nodes of the graph G on the n -node motif M , and we are assuming that the \mathbf{f}_i 's do not contain an explicit dependence on time. The issue of the stability of the steady states of Eqs. (1), i.e. the sets of values $(\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_n^*)$ such that $\dot{\mathbf{x}}_1^* = \dot{\mathbf{x}}_2^* = \dots = \dot{\mathbf{x}}_n^* = 0$, has been investigated in Ref. [11]. Here we focus on the stability of the synchronized dynamics of Eqs. (1), which can be treated analytically within the context of the so-called *Master Stability Function* approach [3, 12–14]. In particular, we restrict to the case in which the equations of motion can be written as:

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) + \sigma \sum_{j=1}^n a_{ij} [\mathbf{H}_{ij}(\mathbf{x}_j) - \mathbf{H}_{ii}(\mathbf{x}_i)] \quad i = 1, \dots, n. \quad (2)$$

where $\mathbf{F}_i(\mathbf{x}) : R^m \rightarrow R^m$ is the function governing the local dynamics of node i , $\mathbf{H}_{ij}(\mathbf{x}_j) : R^m \rightarrow R^m$ describes the influence of node j on node i , $\sigma > 0$ is the coupling strength, and a_{ij} are the elements of the $n \times n$ adjacency matrix of graph M . In the case of a undirected n -node motif M , $a_{ij} = a_{ji} = 1$ iff there is an edge joining node i and node j , and $a_{ij} = a_{ji} = 0$ otherwise. In the case of a directed motif, we assume $a_{ij} = 1$ iff there is a directed edge from node j to node i , while $a_{ij} = 0$ otherwise. Equations (2) can be rewritten as:

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) - \sigma \sum_{j=1}^n l_{ij} \mathbf{H}_{ij}(\mathbf{x}_j) \quad i = 1, \dots, n. \quad (3)$$

where $l_{ij} = \delta_{ij} (\sum_l a_{il}) - a_{ij}$ are the elements of a zero row-sum ($\sum_j l_{ij} = 0 \forall i$) $n \times n$ matrix L with strictly positive diagonal terms ($l_{ii} > 0 \forall i$). In the case of a undirected motif M , L is symmetric and coincides with the standard *Laplacian matrix* of the graph M [3]. In the case of a directed graph, the off-diagonal elements l_{ij} of L are respectively equal to $-a_{ij}$, while the i -th diagonal entry is equal to the in-degree of node i , $k_i^{in} = \sum_l a_{il}$. In order to proceed with the analytic treatment, we make the explicit assumption that the network is made of n identical and identically coupled dynamical systems. This corresponds to take in Eqs. (2) and Eqs. (3) $\mathbf{F}_i(\mathbf{x}_i) \equiv \mathbf{F}(\mathbf{x}) \forall i$, and $\mathbf{H}_{ij}(\mathbf{x}_j) \equiv \mathbf{H}(\mathbf{x}) \forall i, j$. This assumption and the fact that L is zero-row sum, ensure the existence of an invariant set $\mathbf{x}_1(t) = \mathbf{x}_2(t) = \dots = \mathbf{x}_n(t) \equiv \mathbf{x}_s(t)$, representing the complete synchronization manifold \mathcal{S} . The main idea, first proposed by

Pecora and Carrol [12], is that the linear stability analysis of the synchronized state of Eqs. (3) can be divided into a topological and a dynamical part [12, 14]. Since the coupling term of Eqs. (3) vanishes exactly on \mathcal{S} , a necessary condition for the stability of the synchronous state is that the set of $(n - 1) * m$ Lyapunov exponents corresponding to phase space directions transverse to the synchronization manifold are entirely made of negative values. Considering, then, the $m \times n$ column vectors $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ and $\delta\mathbf{X} = (\delta\mathbf{x}_1, \dots, \delta\mathbf{x}_n)^T$, where $\delta\mathbf{x}_i(t) = \mathbf{x}_i(t) - \mathbf{x}_s(t)$ is the deviation of the i^{th} vector state from the synchronization manifold, one gets the variational equation:

$$\delta\dot{\mathbf{X}} = [I_n \otimes \mathbf{J}\mathbf{F}(\mathbf{x}_s) - \sigma L \otimes \mathbf{J}\mathbf{H}(\mathbf{x}_s)] \delta\mathbf{X}, \quad (4)$$

where I_n is the $n \times n$ identity matrix, \otimes stands for the direct product between matrices, and \mathbf{J} denotes the Jacobian operator. The first term in Eq. (4) is block diagonal with $m \times m$ blocks, while the second term can be treated by diagonalizing L .

We first concentrate on the case of **undirected motifs**, i.e. on symmetric and thus diagonalizable laplacian L . Let λ_i be the set of n real eigenvalues of L ($L\mathbf{v}_i = \lambda_i\mathbf{v}_i$, $i = 1, \dots, n$), and \mathbf{v}_i the associated orthonormal eigenvectors ($\mathbf{v}_j^T \cdot \mathbf{v}_i = \delta_{ij}$). If L is symmetric, all its eigenvalues are real, and they can be ordered by size as: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. The arbitrary state $\delta\mathbf{X}$ can be written as $\delta\mathbf{X} = \sum_{i=1}^n \mathbf{v}_i \otimes \zeta_i(t)$, where $\zeta_i \equiv (\zeta_{1,i}, \dots, \zeta_{m,i})$. By substituting into Eq. (4), and using the condition that the eigenvectors are linearly independent, one is finally left with a block diagonalized variational equation, with each of the n blocks having the form of a variational equation for the coefficient $\zeta_k(t)$:

$$\frac{d\zeta_k}{dt} = \mathbf{K}_k \zeta_k, \quad k = 1, \dots, n \quad (5)$$

where $\mathbf{K}_k = [\mathbf{J}\mathbf{F}(\mathbf{x}_s) - \sigma\lambda_k\mathbf{J}\mathbf{H}(\mathbf{x}_s)]$ is the evolution kernel. Each equation in (5) corresponds to a set of m conditional Lyapunov exponents along the eigenmode corresponding to the specific eigenvalue λ_k . For $k = 1$, $\lambda_1 = 0$, and we have the variational equation for the synchronized manifold \mathcal{S} . The m corresponding conditional Lyapunov exponents equal those of the single uncoupled system $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$, therefore no conditions on them will be imposed (in principle, the synchronized state itself can well have positive Lyapunov exponents and be chaotic).

Notice that the Jacobian $\mathbf{J}\mathbf{F}(\mathbf{x}_s)$ and $\mathbf{J}\mathbf{H}(\mathbf{x}_s)$ are the same for each block k , since they are evaluated on the synchronized state. Consequently, the form of each of the blocks in Eqs. (5) is the same, with the only difference being in the multiplier λ_k . This leads one to replace $\sigma\lambda_k$ by ν in Eq. (5), and to consider the generic m -dimensional variational equation:

$$\dot{\zeta} = \mathbf{K}_\nu \zeta = [\mathbf{J}\mathbf{F}(\mathbf{x}_s) - \nu\mathbf{J}\mathbf{H}(\mathbf{x}_s)] \zeta, \quad (6)$$

from which one can extract the set of m conditional Lyapunov exponents as a function of the real parameter $\nu \geq 0$. The parametrical behavior of the largest of such exponents, $\Lambda(\nu)$, is called *Master Stability Function* [12–14]. In fact, given a coupling strength σ , one can locate the point $\sigma\lambda_k$ on the positive ν axis, and the sign of Λ at that point will reveal the stability of that eigenmode. If $\Lambda(\sigma\lambda_k) < 0 \ \forall k = 2, \dots, n$, then the synchronous state is stable at the coupling strength σ .

In order to evaluate whether the stability of the synchronous state is favoured by the topology in a given n -node graph more than in another, we adopt the following measures of stability. First, we assume that $\Lambda(\nu = 0) > 0$, meaning that the uncoupled systems $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ support a chaotic dynamics. For $\nu > 0$, there are three possible behaviors of $\Lambda(\nu)$, defining

three possible classes for the choice of the functions $\mathbf{F}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$. Case I (II) corresponds to a monotonically increasing (decreasing) $\Lambda(\nu)$. Case III admits negative values of $\Lambda(\nu)$ in the range $\nu_{c1} < \nu < \nu_{c2}$ (see Fig 5.1 of Ref. [3]). For systems in class I, one can never stabilize synchronization in any graph topology. In fact, for any σ and any eigenvalues' distributions, the product $\sigma\lambda_k$ always leads to a positive maximum Lyapunov exponent, and therefore the synchronization manifold \mathcal{S} is always transversally unstable. Class II systems always admit synchronization for a large enough σ . In fact, given any eigenvalue distributions (any graph topology) it is sufficient to select $\sigma > \nu_c/\lambda_2$ ($\lambda_2 \neq 0$ in a connected graph [15]) to warrant that all transverse directions to \mathcal{S} have associated negative Lyapunov exponents. The synchronous state will be stable for smaller values of σ in a graph with a larger λ_2 , so that λ_2 can be used as a measure of the stability of the synchronous state (SSS). For systems in class III, the stability condition is satisfied when $\sigma > \nu_{c1}/\lambda_2$ and $\sigma < \nu_{c2}/\lambda_N$, indicating that the more packed the eigenvalues of L are, the higher is the chance of having all Lyapunov exponents into the stability range [14]. Consequently, the ratio λ_2/λ_n can be used as a measure of SSS. Classes II and III include a large number of functions F , describing several relevant dynamical systems, as the Lorenz and Rössler chaotic oscillators, and the Chua oscillator. It is important to notice that not only F , but also H has a role in determining to which class a specific dynamical system belongs to. As an example, a nearest neighbor diffusive coupling on the Rössler chaotic system yields a class II (class III) Master Stability Function, when the function H extracts the second (the first) component of the vector field [16]. In Fig. 1 (panel a and b) we report the two indices of SSS, namely λ_2 (class II) and λ_2/λ_4 (class III), for the six 4-node undirected motifs. We observe a general increase in the SSS's as the number of the edges in the motif increases. Such an increase in SSS is in agreement with the decrease of the synchronization threshold observed numerically in the Kuramoto model by Moreno et al. [17]. The two measures of SSS we propose are also in good agreement with the natural conservation ratio (NCR) for the same 4-node motifs in the yeast protein interaction network reported in panel c). The NCR is a measure proposed in Ref. [18] to quantify the conservation of a given motif in the evolution across species, and is highly correlated to the motif Z-score. In panel d) and e) we show that SSS's and NCR are linearly correlated: we have obtained a correlation coefficient respectively equal to 0.94 and 0.93. This is an indication that motifs displaying an improved stability of cooperative activities (as synchronous states) are preserved across evolution with a higher probability.

We now turn our attention to **directed motifs**. In a directed graph, the matrix L is asymmetric and in general not always diagonalizable. Nevertheless, L can be transformed into a Jordan canonical form, and it has been proven that the same condition valid for diagonalizable networks ($\Lambda(\sigma\lambda_k) < 0 \ \forall k = 2, \dots, n$) also applies to non-diagonalizable networks [19]. In addition, the spectrum of L is either real or made of pairs of complex conjugates. Because of the zero row-sum condition, L always admits $\lambda_1 = 0$, and the other eigenvalues $\lambda_k = \lambda_k^R + i\lambda_k^I$, $k = 2, \dots, n$ (having non negative real parts according to the Gerschgorin's circle theorem [20]) can be ordered by increasing real part ($0 \leq \lambda_2^R \leq \dots \leq \lambda_n^R$). Consequently, the parametric equation (6) has to be studied for complex values of the parameter $\nu = \nu^R + i\nu^I$. This yields a master stability function $\Lambda(\nu)$ as a surface over the complex plane ν , that generalizes the plots for the case ν real. By calling \mathcal{R} the region in the complex plane where $\Lambda(\nu)$ provides a negative Lyapunov exponent, the stability condition for the synchronous state is that the set $\{\sigma\lambda_k, k = 2, \dots, n\}$ be entirely contained in \mathcal{R} for a given σ . This is best accomplished for connection topologies that make λ_2^R as large as possible for class I systems, and for topologies that simultaneously make $\frac{\lambda_2^R}{\lambda_n^R}$ as large as possible and $\max_{k \geq 2} \{|\lambda_k^I|\}$ as small as possible, for class II systems.

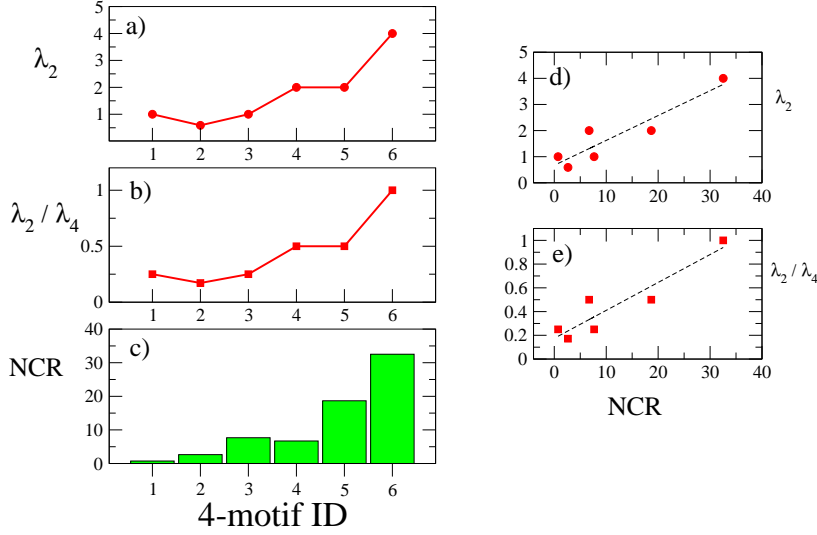


Fig. 1 – The value of SSS for each of the six 4-node undirected motifs are reported in panel a) for class II systems and in panel b) for class III, and compared with the natural conservation rates (NCR) in the yeast protein interaction network [18]. The motif identification number is the same as in Ref. [7]. In panel d) and e) we plot the values of SSS as function of the NCR (symbols), and the linear fittings obtained (dashed lines).

In Fig. 2, we consider the thirteen 3-node directed motifs. Two of them, namely motifs #3 and motif #11 give rise to non-diagonalizable L . Motif #8 is the only case where the eigenvalues are not real. In the left (right) panels we report λ_2^R for class II systems ($\lambda_2^R / \lambda_3^R$ for class III systems). The SSS measures are compared with the Z-score profile obtained for five different real biological networks, and shown as histograms in the figure. Both class I and class II systems exhibit an average increase of SSS as a function of the number of links in the motif. However, the overall agreement of the SSS and the Z-score profiles is not as good as in the case of undirected 4-motifs. Here, we have obtained rather small values (ranging from 0.1 to 0.3) of the correlation coefficient, with a better agreement found in the case of the STKE network (panels c), Drosophila (panels d) and C.elegans (panels e), rather than in the transcriptional regulatory networks (panels a and b). This might be due to the fact that synchronization processes are more important in neural systems than in other biological systems as transcriptional networks, especially the simplest ones (*E. coli* and *S. cerevisiae*). We have also reported in figure, as dashed lines, the measure of the stability of stationary states proposed by Prill et al. [11]. Such a measure seems to be better indicated for those systems where the stability of stationary states can be a more relevant dynamical quantity to investigate than the stability of synchronous states. Fig. 2 clearly indicates that in some motifs, SSS and Z-score are better correlated than in others. Hence, for each motif M , we have defined an *overlap coefficient* O_M as: $O_M = SSS_M \times Z_M$. The maximum possible value $O_M = 1$ indicates a perfect correlation between SSS and Z-score. The overlap coefficients obtained for the five studied systems are reported in Fig. 3. For both class II and class III systems we have high values of the overlap for motifs: 1, 7, 10, 12.

Finally, we have considered the 199 4-node directed motifs. Here we report the results for three of the most statistically relevant motifs found in biological networks: the bifan, the biparallel and the feedback loop (see Ref. [5]). Such three motifs correspond all to cases in

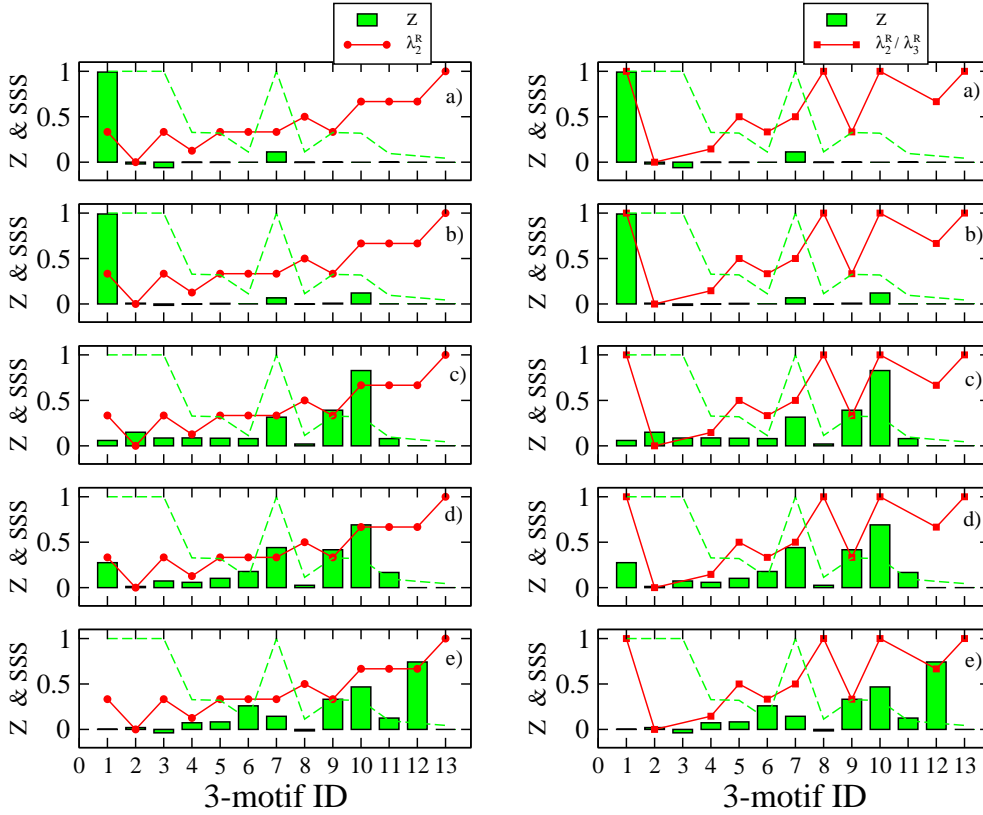


Fig. 2 – The SSS of each of the thirteen 3-node directed motifs is reported (continuous line) for class II (left panels), and class III systems (right panels). For class III system the SSS values has been normalized to the maximum value so to vary in the range $[0,1]$. The SSS values are compared with the Z-score (hystograms) and with a measure of the stability of stationary states (dashed line) from Ref. [11], in five different biological networks: the transcriptional regulatory networks of *E. coli* (panels a) and *S. cerevisiae* (panel b), the signal transduction knowledge environment (STKE) network (panel c), the developmental transcriptional network of *Drosophila melanogaster* (panel d), and the neural connection map of *C. elegans* (panel e).

which L can be diagonalized. The biparallel graph, that is abundant in the *C. elegans* and in transcriptional networks, has real eigenvalues and a relatively high value of SSSs: $\lambda_2^R = 1$ and $\lambda_2^R/\lambda_4^R = 0.5$. The same is true for the 4-node feedback loop (also found abundant in electric circuits [5]), having $\lambda_2^R = 1$, $\lambda_2^R/\lambda_4^R = 0.5$ and $\max_{k \geq 2} \{|\lambda_k^I|\} = 1$. Conversely, the bifan is not compatible with synchronization for any choice of $\mathbf{F}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$, and for any value of σ , since $\lambda_2 = 0$ and we have assumed the case of networked chaotic systems ($\Lambda(\nu = 0) > 0$). In fact, $\lambda_2^R \neq 0$ iff the graph embeds an oriented spanning tree, (i.e., there is a node from which all other nodes can be reached by following directed links) [19,21] and this condition, that generalizes the notion of connectedness for undirected graphs [15] to directed graphs, is not valid in the case of the bifan.

We warmly thank R.J. Prill and A. Levchenko for having provided us with their results on the stability of stationary states, and G. Russo for useful comments. S.B. acknowledges the Yeshaya Horowitz Association through the Center for Complexity Science.

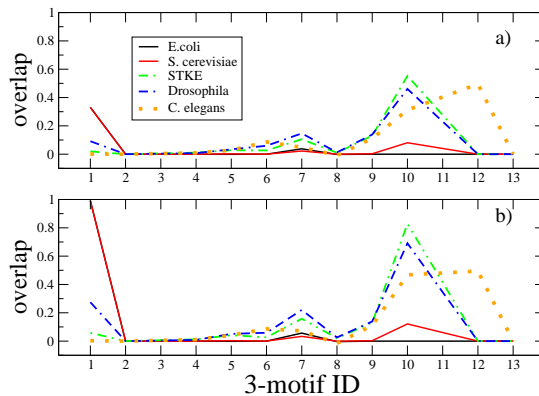


Fig. 3 – The overlap coefficients for each of the thirteen 3-node directed motifs, and the five biological networks considered, are reported for class II (panel a) and class III systems (panel b).

REFERENCES

- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [2] M.E.J. Newman, *SIAM Review* **45**, 167 (2003).
- [3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006).
- [4] S. Shen-Orr, R. Milo, S. Mangan and U. Alon, *Nature Genetics* **31**, 64 (2002).
- [5] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [6] S. Mangan and U. Alon, *Proc Natl Acad Sci USA* **100**, 11980 (2003).
- [7] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, *Science* **303**, 1538 (2004).
- [8] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, *Bioinformatics* **20**, 1746 (2004).
- [9] S. Valverde and R. V. Solé, *Phys. Rev.* **E72**, 026107 (2005).
- [10] A. Vázquez et al., *PNAS* **101**, 17940 (2004).
- [11] R.J. Prill, P.A. Iglesias and A. Levchenko, *PLoS Biology* **3**, 1881 (2005).
- [12] L.M. Pecora and T.L. Carroll, *Phys. Rev. Lett.* **80**, 2109 (1998).
- [13] K.S. Fink, G. Johnson, T.L. Carroll, D. Mar and L.M. Pecora, *Phys. Rev.* **E61**, 5080 (2000).
- [14] M. Barahona and L.M. Pecora, *Phys. Rev. Lett.* **89**, 054101 (2002).
- [15] M. Fiedler, *Czech. Math. J.* **23** (1973) 298.
- [16] D.-U. Hwang et al., *Phys. Rev. Lett.* **94**, 138701 (2005). M. Chavez et al, *Phys. Rev. Lett.* **94**, 218701 (2005).
- [17] Y. Moreno, M. Vázquez-Prada, and A. F. Pacheco, *Physica A* **343**, 279 (2004).
- [18] S. Wuchty, Z. N. Oltvai, and A.-L. Barabási, *Nat. Genet.* **35**, 176-179 (2003).
- [19] T. Nishikawa and A. E. Motter, *Phys. Rev.* **E73**, 065106R (2006).
- [20] S. A. Gershgorin, *Izv. Akad. Nauk. SSSR, Ser. Mat.* **7**, 749 (1931).
- [21] C. W. Wu, *Linear Algebr. Appl.* **402**, 207 (2005).